

# Microsoft Certified: Azure Data Scientist Associate – Skills Measured

**This document contains the skills measured on the exams associated with this certification. It does not include any upcoming or recent changes that have been made to those skills. For more information about upcoming or recent changes, see the associated exam details page(s).**

## Exam DP-100: Designing and Implementing a Data Science Solution on Azure

### Skills Measured

NOTE: The bullets that appear below each of the skills measured are intended to illustrate how we are assessing that skill. This list is not definitive or exhaustive.

NOTE: In most cases, exams do NOT cover preview features, and some features will only be added to an exam when they are GA (General Availability).

### Set up an Azure Machine Learning Workspace (30-35%)

#### Create an Azure Machine Learning workspace

- create an Azure Machine Learning workspace
- configure workspace settings
- manage a workspace by using Azure Machine Learning studio

#### Manage data objects in an Azure Machine Learning workspace

- register and maintain data stores
- create and manage datasets

#### Manage experiment compute contexts

- create a compute instance
- determine appropriate compute specifications for a training workload
- create compute targets for experiments and training

### Run Experiments and Train Models (25-30%)

#### Create models by using Azure Machine Learning Designer

- create a training pipeline by using Azure Machine Learning designer
- ingest data in a designer pipeline
- use designer modules to define a pipeline data flow
- use custom code modules in designer

### **Run training scripts in an Azure Machine Learning workspace**

- create and run an experiment by using the Azure Machine Learning SDK
- consume data from a data store in an experiment by using the Azure Machine Learning SDK
- consume data from a dataset in an experiment by using the Azure Machine Learning SDK
- choose an estimator for a training experiment

### **Generate metrics from an experiment run**

- log metrics from an experiment run
- retrieve and view experiment outputs
- use logs to troubleshoot experiment run errors

### **Automate the model training process**

- create a pipeline by using the SDK
- pass data between steps in a pipeline
- run a pipeline
- monitor pipeline runs

## **Optimize and Manage Models (20-25%)**

### **Use Automated ML to create optimal models**

- use the Automated ML interface in Azure Machine Learning studio
- use Automated ML from the Azure Machine Learning SDK
- select scaling functions and pre-processing options
- determine algorithms to be searched
- define a primary metric
- get data for an Automated ML run
- retrieve the best model

### **Use Hyperdrive to tune hyperparameters**

- select a sampling method
- define the search space
- define the primary metric

- define early termination options
- find the model that has optimal hyperparameter values

### **Use model explainers to interpret models**

- select a model interpreter
- generate feature importance data

### **Manage models**

- register a trained model
- monitor model history
- monitor data drift

## **Deploy and Consume Models (20-25%)**

### **Create production compute targets**

- consider security for deployed services
- evaluate compute options for deployment

### **Deploy a model as a service**

- configure deployment settings
- consume a deployed service
- troubleshoot deployment container issues

### **Create a pipeline for batch inferencing**

- publish a batch inferencing pipeline
- run a batch inferencing pipeline and obtain outputs

### **Publish a designer pipeline as a web service**

- create a target compute resource
- configure an Inference pipeline
- consume a deployed endpoint